Supporting Information

Understanding the Sequence Preference of Recurrent RNA Building Blocks using Quantum Chemistry:

The Intrastrand RNA Dinucleotide Platform

Arnošt Mládek, Judit E. Šponer, Petr Kulhánek, Xiang-Jun Lu, Wilma K. Olson, and Jiří Šponer**

Table S1 Crystal structures used to derive GpU, ApA, and UpC dinucleotide platforms. In two cases (1q9a and 483d), both alternative atom positions of the 5′-residue were used (labeled as A/B).

Platform type	PDB ID	Residues:Chain	Label	X-ray structure resolution [Å]	Reference
GpU	1q9a	2655-2656:A	GU-1q9a-1A	1.04	[S1]
		2655-2656:A	GU-1q9a-1B		
	483d	2655-2656:A	GU-483d-1A	1.11	[S2]
		2655-2656:A	GU-483d-1B		
	480d	2655-2656:A	GU-480d-1	1.50	[S2]
	1msy	2655-2656:A	GU-1msy-1	1.41	[S3]
	3dw4	2655-2656:A	GU-3dw4-1	0.97	[S4]
	3dw6	2655-2656:A	GU-3dw6-1	1.00	[S4]
	3dvz	2655-2656:A	GU-3dvz-1	1.00	[S4]
	1q96	10-11:A	GU-1q96-1	1.75	[S1]
		10-11:B	GU-1q96-2		
		10-11:C	GU-1q96-3		
	1q93	10-11:A	GU-1q93-1	2.25	[S1]
		10-11:C	GU-1q93-2		
	1y27	62-31:X	GU-1y27-1	2.40	[S5]
	2ees	62-31:A	GU-2ees-1	1.75	[S6]
	1u8d	62-31:A	GU-1u8d-1	1.95	[S7]
	2b57	62-31:A	GU-2b57-1	2.15	[S8]
	2g9c	62-31:A	GU-2g9c-1	1.70	[S9]
	2qus	20-21:A	GU-2qus-1	2.40	[S10]
	2quw	20-21:A	GU-2quw-1	2.20	[S10]
	1jj2	1292-1293:0	GU-1jj2-1	2.40	[S11]
		175-176:0	GU-1jj2-2		
		213-214:0	GU-1jj2-3		
		358-359:0	GU-1jj2-4		
		381-382:0	GU-1jj2-5		
		464-465:0	GU-1jj2-6		
		1370-1371:0	GU-1jj2-7		
		1971-1972:0	GU-1jj2-8		
		2692-2693:0	GU-1jj2-9		
		78-79:9	GU-1jj2-10		
	3dil	27-28:A	GU-3dil-1	1.90	[S12]
ApA	1gid	218-219:A	AA-1gid-1	2.50	[S13]
		218-219:B	AA-1gid-3		
		225-226:A	AA-1gid-2		
		225-226:B	AA-1gid-4		
	1hr2	225-226:A	AA-1hr2-1	2.25	[S14]
		225-226:B	AA-1hr2-2		
		171-172:A	AA-1hr2-3		
		171-172:B	AA-1hr2-4		
		218-219:B	AA-1hr2-5		
	2r8s	172-172:R	AA-2r8s-1	1.95	[S15]
		225-226:R	AA-2r8s-2		
	1jj2	1193-1194:0	AA-1jj2-1	2.40	[S11]
		441-442:0	AA-1jj2-2		
		51-52:9	AA-1jj2-3		
UpC	1drz	155-156:B	UC-1drz-1	2.30	[S16]
	1jj2	1009-1010:0	UC-1jj2-1	2.40	[S11]

	1sj3	155-156:R	UC-1sj3-1	2.20	[S17]
	1u0b	73-74:A	UC-1u0b-1	2.30	[S18]
	1vc7	155-156:B	UC-1vc7-1	2.45	[S17]

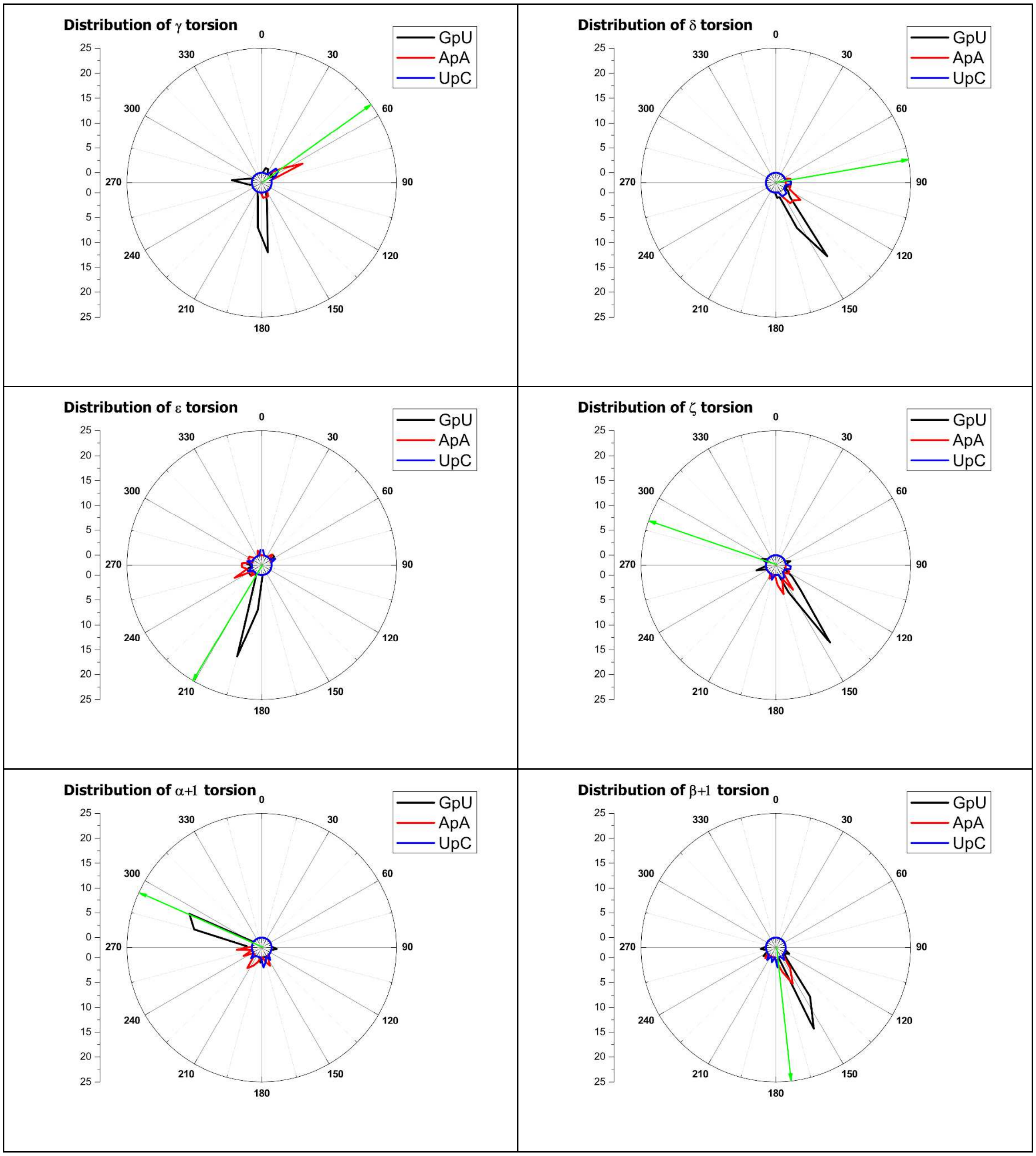
References

[S1] Correll, C. C.; Beneken, J.; Plantinga, M. J.; Lubbers, M.; Chan, Y. L. *Nucleic Acids Res*. **2003**, *31*, 6806-6818
[S2] Correll, C. C.; Wool, I. G.; Munishkin, A. *J. Mol. Biol*. **1999**, *292*, 275-287
[S3] Correll, C. C.; Swinger, K. *RNA* **2003**, *9*, 355-363
[S4] Olieric, V.; Rieder, U.; Lang, K.; Serganov, A.; Schulze-Briese, C.; Micura, R.; Dumas, P.; Ennifar, E. *RNA* **2009**, *15*, 707-715
[S5] Serganov, A.; Yuan, Y. R.; Pikovskaya, O.; Polonskaia, A.; Malinina, L.; Phan, A. T.; Hobartner, C.; Micura, R.; Breaker, R. R.; Patel, D. *J. Chem. Biol*. **2004**, *11*, 1729-1741
[S6] Gilbert, S. D.; Love, C. E.; Edwards, A. L.; Batey, R. T. *Biochemistry* **2007**, *46*, 13297-13309
[S7] Batey, R. T.; Gilbert, S. D.; Montange, R. K. *Nature* **2004**, *432*, 411-415
[S8] Gilbert, S. D.; Stoddard, C. D.; Wise, S. J.; Batey, R. T. *J. Mol. Biol*. **2006**, *359*, 754-768
[S9] Gilbert, S. D.; Mediatore, S. J.; Batey, R. T. *J. Am. Chem. Soc*. **2006**, *128*, 14214-14215
[S10] Chi, Y. I.; Martick, M.; Lares, M.; Kim, R.; Scott, W. G.; Kim, S. H. *PLoS Biol*. **2008**, *6*, 2060-2068
[S11] Klein, D. J.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. *EMBO J*. **2001**, *20*, 4214-4221
[S12] Serganov, A.; Huang, L.; Patel, D. J. *Nature* **2008**, *455*, 1263-1276
[S13] Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678-1685
[S14] Juneau, K.; Podell, E.; Harrington, D. J.; Cech, T. R. *Structure* **2001**, *9*, 221-231
[S15] Ye, J. D.; Tereshko, V.; Frederiksen, J. K.; Koide, A.; Fellouse, F. A.; Sidhu, S. S.; Koide, S.; Kossiakoff, A. A.; Piccirilli, J. A. *Proc. Natl. Acad. Sci. U. S. A* **2008**, *105*, 82-87
[S16] Ferre-D'Amare, A. R.; Zhou, K.; Doudna, J. A. *Nature* **1998**, *395*, 567-574
[S17] Ke, A.; Zhou, K.; Ding, F.; Cate, J. H.; Doudna, J. A. *Nature* **2004**, 429, 201-205
[S18] Hauenstein, S.; Zhang, C. M.; Hou, Y. M.; Perona, J. J. *Nat. Struct. Mol. Biol*. **2004**, *11*, 1134-1141

Table S2 Backbone and glycosidic torsion angles, in degrees, of the studied dinucleotide platforms supplemented with polar distribution plots. The green arrow in each graph denotes the average torsion of the A-RNA reference state.

System	γ	δ	ε	ζ	α+1	β+1	γ+1	δ+1	χ	χ+1
GU-1q9a-1A	179	147	192	142	291	152	39	89	263	183
GU-1q9a-1B	275	170	194	128	286	152	39	89	287	183
GU-483d-1A	178	147	194	142	288	152	42	90	261	183
GU-483d-1B	275	164	189	134	288	152	42	90	284	183
GU-480d-1	267	158	190	140	295	146	37	88	277	181
GU-1msy-1	179	149	193	141	287	158	38	92	266	186
GU-3dw4-1	183	152	192	142	286	157	40	90	272	185
GU-3dw6-1	184	151	193	141	285	157	40	90	274	184
GU-3dvz-1	181	152	193	141	290	153	39	90	269	182
GU-1q96-1	175	147	195	141	296	148	34	85	258	185
GU-1q96-2	271	151	187	145	297	150	32	86	271	187
GU-1q96-3	175	144	200	139	299	151	30	87	255	189
GU-1q93-1	174	150	193	136	295	146	38	89	263	187
GU-1q93-2	173	149	194	141	296	145	39	86	256	184
GU-1y27-1	162	100	43	295	226	268	85	100	252	209
GU-2ees-1	179	148	249	124	263	178	34	102	268	219
GU-1u8d-1	190	139	235	142	286	153	27	90	267	209
GU-2b57-1	183	137	251	146	264	154	32	83	268	204
GU-2g9c-1	185	143	226	137	288	155	38	146	263	210
GU-2qus-1	24	93	53	257	96	112	336	166	266	238
GU-2quw-1	18	123	278	72	180	234	89	129	278	217
GU-1jj2-1	63	88	183	256	288	194	50	87	194	186
GU-1jj2-2	174	148	196	143	287	153	42	86	261	185
GU-1jj2-3	182	148	185	154	293	149	42	83	264	183
GU-1jj2-4	274	149	185	158	276	149	53	85	281	189
GU-1jj2-5	280	156	190	140	290	137	46	82	275	185
GU-1jj2-6	177	150	195	145	290	151	44	85	258	182
GU-1jj2-7	164	149	192	145	291	149	41	83	261	188
GU-1jj2-8	43	142	197	134	288	145	39	84	276	189
GU-1jj2-9	179	148	189	150	294	141	41	82	255	185
GU-1jj2-10	178	148	185	150	291	149	44	85	268	185
GU-3dil-1	181	141	196	149	288	153	39	85	262	187
AA-1gid-1	64	140	243	168	260	144	39	86	233	179
AA-1gid-3	60	138	243	168	262	144	38	86	233	178
AA-1gid-2	50	123	241	168	245	150	47	89	220	184
AA-1gid-4	50	122	244	168	243	151	47	89	222	187
AA-1hr2-1	48	133	275	149	212	155	50	86	249	187
AA-1hr2-2	60	115	291	145	200	154	55	87	226	189
AA-1hr2-3	156	129	308	172	196	166	43	85	194	181
AA-1hr2-4	64	109	345	129	165	175	55	86	206	186
AA-1hr2-5	177	79	49	112	159	164	61	85	175	195
AA-2r8s-1	60	137	265	148	212	153	57	84	237	185
AA-2r8s-2	60	125	278	144	215	155	47	81	232	189
AA-1jj2-1	58	149	262	206	205	167	54	83	210	180
AA-1jj2-2	61	149	225	179	263	147	43	82	234	188
AA-1jj2-3	165	156	257	153	157	225	75	82	219	197
UC-1drz-1	55	130	354	149	174	149	43	80	240	188
UC-1jj2-1	66	151	249	196	221	172	45	77	231	203

UC-1sj3-1	48	144	280	150	162	216	68	87	243	196
UC-1u0b-1	56	84	62	105	147	173	56	81	188	214
UC-1vc7-1	47	92	7	95	172	191	49	81	215	183



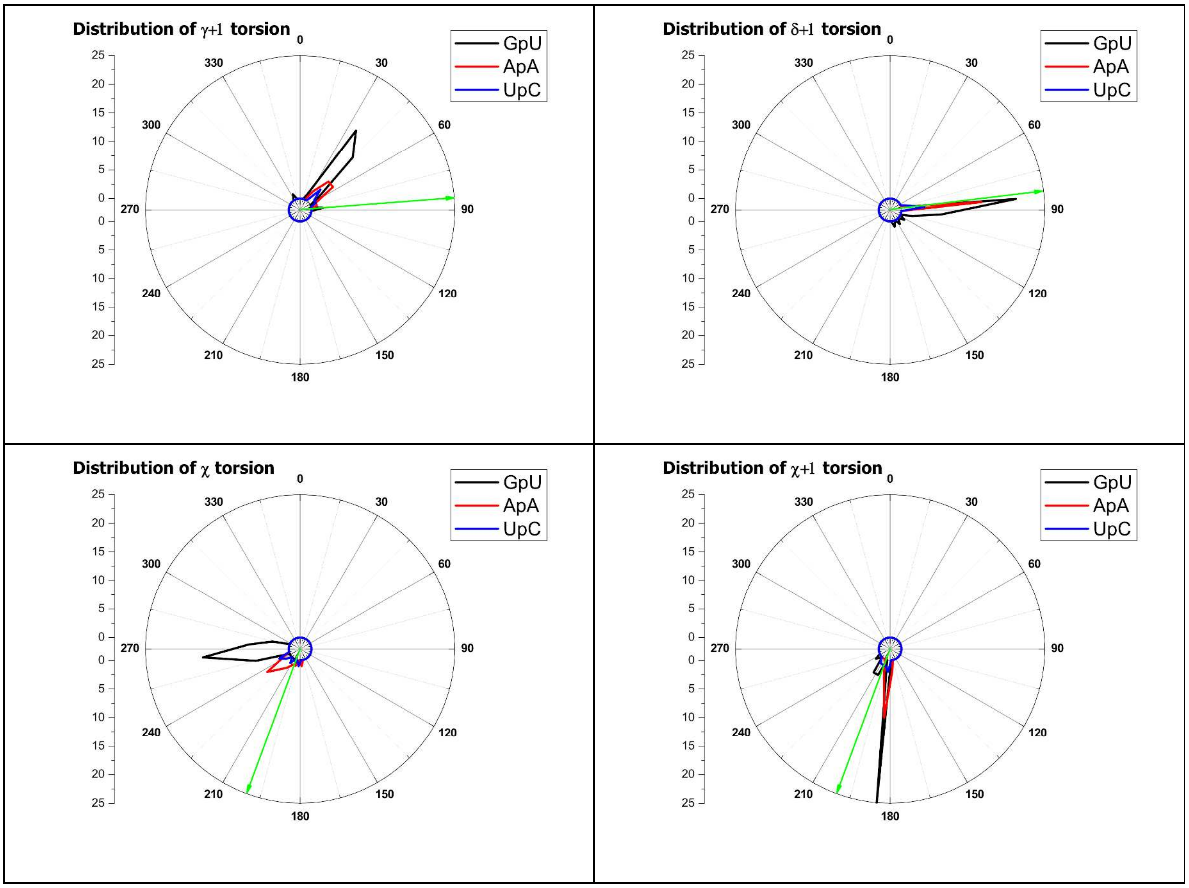

Distribution of γ+1 torsion
Distribution of δ+1 torsion
Distribution of χ torsion
Distribution of χ+1 torsion
GpU
ApA
UpC
0
30
60
90
120
150
180
210
240
270
300
330
25
20
15
10
5
0

Table S3 DF-MP2/aug-cc-pVDZ CP-corrected (i.e. BSSE-free) interaction energies of GU, AA, and UC base pairs. The pairs were derived from the TPSS-D/LP optimized platform dinucleotides. The appended methyl groups were relaxed using M06/6-31+G(d, p) prior to computation of the interaction energies.

Base pair	System	ΔE [kcal.mol^{-1}]
GU	GU-1q9a-1A	–8.1
	GU-1q9a-1B	–8.2
	GU-483d-1A	–8.1
	GU-483d-1B	–8.3
	GU-480d-1	–8.2
	GU-1msy-1	–8.1
	GU-3dw4-1	–8.0
	GU-3dw6-1	–8.0
	GU-3dvz-1	–8.3
	GU-1q96-1	–8.1
	GU-1q96-2	–8.1
	GU-1q96-3	–7.5
	GU-1q93-1	–7.6
	GU-1q93-2	–8.1
	GU-1y27-1	–8.8
	GU-2ees-1	–6.5
	GU-1u8d-1	–6.1
	GU-2b57-1	–7.8
	GU-2g9c-1	–8.5
	GU-2qus-1	–8.9
	GU-2quw-1	–8.2
	GU-1jj2-1	–8.1
	GU-1jj2-2	–8.1
	GU-1jj2-3	–8.1
	GU-1jj2-4	–7.9
	GU-1jj2-5	–8.1
	GU-1jj2-6	–8.1
	GU-1jj2-7	–8.0
	GU-1jj2-8	–7.7
	GU-1jj2-9	–8.0
	GU-1jj2-10	–7.9
	GU-3dil-1	–7.7
AA	AA-1gid-1	–6.5
	AA-1gid-3	–6.5
	AA-1gid-2	–7.1
	AA-1gid-4	–7.1
	AA-1hr2-1	–6.6
	AA-1hr2-2	–7.1
	AA-1hr2-3	–7.3
	AA-1hr2-4	–6.9
	AA-1hr2-5	–2.6
	AA-2r8s-1	–6.3
	AA-2r8s-2	–6.4
	AA-1jj2-1	–6.9
	AA-1jj2-2	–6.4
	AA-1jj2-3	–7.1
UC	UC-1drz-1	–5.6

	UC-1jj2-1	–5.6
	UC-1sj3-1	–5.5
	UC-1u0b-1	–4.8
	UC-1vc7-1	–4.9

Table S4 RI-MP2/CBS relative energies of rSPSOM models of GpU, ApA, and UpC dinucleotide platforms. The energies are expressed as relative values with respect to the rSPSOM model of the A-RNA reference state.

rSPSOM type	System	E [kcal.mol^{-1}]
GpU	GU-1q9a-1A	–0.8
	GU-1q9a-1B	+5.0
	GU-483d-1A	–0.9
	GU-483d-1B	+4.7
	GU-480d-1	+1.2
	GU-1msy-1	–0.9
	GU-3dw4-1	–0.6
	GU-3dw6-1	–0.3
	GU-3dvz-1	–0.3
	GU-1q96-1	–0.8
	GU-1q96-2	+0.4
	GU-1q96-3	–1.2
	GU-1q93-1	+0.5
	GU-1q93-2	–0.7
	GU-1y27-1	+1.4
	GU-2ees-1	+7.0
	GU-1u8d-1	–0.2
	GU-2b57-1	+0.4
	GU-2g9c-1	–0.4
	GU-2qus-1	+4.6
	GU-2quw-1	+4.0
	GU-1jj2-1	–3.6
	GU-1jj2-2	–1.5
	GU-1jj2-3	–1.2
	GU-1jj2-4	–0.4
	GU-1jj2-5	+0.6
	GU-1jj2-6	–1.4
	GU-1jj2-7	–1.2
	GU-1jj2-8	–0.3
	GU-1jj2-9	–1.2
	GU-1jj2-10	–0.6
	GU-3dil-1	–2.2
ApA	AA-1gid-1	+0.9
	AA-1gid-3	+0.4
	AA-1gid-2	+1.2
	AA-1gid-4	+1.1
	AA-1hr2-1	+0.4
	AA-1hr2-2	+2.4
	AA-1hr2-3	+6.2
	AA-1hr2-4	+6.1
	AA-1hr2-5	+3.9
	AA-2r8s-1	+2.1
	AA-2r8s-2	–0.1
	AA-1jj2-1	+1.3
	AA-1jj2-2	–1.1
	AA-1jj2-3	+2.8
UpC	UC-1drz-1	+6.1
	UC-1jj2-1	+2.1

	UC-1sj3-1	–3.1
	UC-1u0b-1	+0.9
	UC-1vc7-1	–0.9

Table S5 Grouping of GpU, ApA, and UpC rSPSOM platform models based on the conformational classification software of Richardson *et al*. Models that could not be assigned a conformational type are marked as unknown (U). Roman numerals denote the categories discussed in the text and symbols in parentheses are those suggested by Richardson *et al.* and reported in the computer output. The suiteness value between 0 and 1 denotes the similarity of the rSPSOM to the average structure in the particular class. The A/B suffixes denoting particular conformational substates of the bi-conformational structures are highlighted in boldface.

rSPSOM type	System	Conformational class	Suiteness
GpU	GU-1q9a-1**A**	II (#a)	0.926
	GU-1q9a-1**B**	II (#a)	0.103
	GU-483d-1**A**	II (#a)	0.923
	GU-483d-1**B**	II (#a)	0.379
	GU-480d-1	II (#a)	0.667
	GU-1msy-1	II (#a)	0.842
	GU-3dw4-1	II (#a)	0.894
	GU-3dw6-1	II (#a)	0.904
	GU-3dvz-1	II (#a)	0.896
	GU-1q96-1	II (#a)	0.847
	GU-1q96-2	II (#a)	0.768
	GU-1q96-3	II (#a)	0.671
	GU-1q93-1	II (#a)	0.862
	GU-1q93-2	II (#a)	0.922
	GU-1y27-1	U (!!)	-
	GU-2ees-1	III (0a)	0.095
	GU-1u8d-1	III (0a)	0.279
	GU-2b57-1	III (0a)	0.213
	GU-2g9c-1	IV (4g)	0.431
	GU-2qus-1	U (!!)	-
	GU-2quw-1	U (!!)	-
	GU-1jj2-1	I (&a)	0.785
	GU-1jj2-2	II (#a)	0.977
	GU-1jj2-3	II (#a)	0.928
	GU-1jj2-4	II (#a)	0.681
	GU-1jj2-5	II (#a)	0.774
	GU-1jj2-6	II (#a)	0.989
	GU-1jj2-7	II (#a)	0.982
	GU-1jj2-8	II (#a)	0.803
	GU-1jj2-9	II (#a)	0.927
	GU-1jj2-10	II (#a)	0.959
	GU-3dil-1	II (#a)	0.832
ApA	AA-1gid-1	III (0a)	0.196
	AA-1gid-3	III (0a)	0.184
	AA-1gid-2	U (!!)	-
	AA-1gid-4	U (!!)	-
	AA-1hr2-1	IV (4g)	0.348
	AA-1hr2-2	U (!!)	-
	AA-1hr2-3	IV (4g)	0.010
	AA-1hr2-4	U (!!)	-
	AA-1hr2-5	U (!!)	-
	AA-2r8s-1	IV (4g)	0.520
	AA-2r8s-2	IV (4g)	0.039
	AA-1jj2-1	IV (4g)	0.146
	AA-1jj2-2	III (0a)	0.124

	AA-1jj2-3	U (!!)	-
UpC	UC-1drz-1	U (!!)	-
	UC-1jj2-1	IV (4g)	0.339
	UC-1sj3-1	U (!!)	-
	UC-1u0b-1	U (!!)	-
	UC-1vc7-1	U (!!)	-

A-RNA optimization constraints

Optimization constraints were imposed on the quasi-β and quasi-ε+1 torsions of the A-RNA backbone based on the following rationale: (i) given that the A-RNA conformational substate is periodic, corresponding torsions in different nucleotide units are assumed to adopt the same values (ignoring any sequence preferences), and (ii) if the quasi-β torsion is free to relax, the 5´-terminal methoxy group is pulled toward the phosphate group via a C_{Met}-H...O1P interaction, with the resultant C_{Met}...O1P distance of ~ 3.5 Å leading to a non-realistic stabilization of the structure. Such extensive bending of the backbone is biologically irrelevant as it would lead to a strong electrostatic repulsion between the negatively charged phosphates. In line with this, no such backbone deviation has been observed in the experimentally available geometries. For similar reasons, the quasi-ε+1 dihedral was fixed as well at the A-RNA average value. Note that no such backbone flexion was observed for the platform systems throughout geometry minimization and thus both quasi-β and quasi-ε+1 dihedrals were optimized.

Clarification of exclusion the AA-1hr2-5 base pair from statistics

The lower stability of the AA-1hr2-5 system is probably due to the inaccurately resolved experimental geometry of the sugar-phosphate backbone as the 3´-nucleotide is shifted in such a way that the C2 of the 5´-adenine and the N6 of the 3´-adenine are positioned as close as 2.7 Å. Despite extensive constraints imposed on the backbone and the two glycosidic torsion angles, the experimentally predicted coplanarity and the pairing pattern of the nucleobases are disrupted upon optimization.

GpU rSPSOM Outliers - In-depth Analysis

The **A/B** suffixes distinguishing conformational substates of the bi-conformational structures, i.e. GU-1q9a-1 and GU-483d-1, are highlighted in boldface.

- **Platforms derived from high-resolution structures**: 1q9a and 483d

Two rSPSOMs were derived from each of the 1q9a and 483d structures (high resolution X-ray structures of the *Escherichia coli* sarcin/ricin domain which included refinement of two 5´-residue geometries), here named GU-1q9a-1**A/B** and GU-483d-1**A/B**, where **A/B** refers to one of the two alternative conformations of the 5´-residue (G). Both **B**-labeled substates show noticeably lower intrinsic stabilities compared to the **A** substates. This difference correlates well with the low *S*-values of the **B**-alternatives (*S*-values of **A/B** conformational variants being 0.93/0.10 and 0.92/0.38 for GU-1q9a-**A/B** and GU-483d-**A/B**, respectively). The rSPSOM model of the GU-1q9a-1**B** backbone, which is 5.8 kcal.mol^{-1} higher in energy than that of the **A** variant (and 5.0 kcal.mol^{-1} above that of the A-RNA reference), contains a steric clash between H2´...H5´ (2.1 Å) brought about by high δ (170°) and a *gauche*$^{-}$ state of γ. The uncompensated shift of ζ, from 142° (**A** variant) to 128° (**B** variant), in combination with high δ, probably weakens the O2´...O2P interaction. Given that the **A/B** alternatives only differ in the sugar-phosphate backbone of the 5´-residue, the structural variations of GU-1q9a-1**B** might represent a destabilization of the system. The rationale for the lower stability of the rSPSOM model of GU-483d-1**B**, with energy 5.6 kcal.mol^{-1} above the **A**-alternative (and 4.7 kcal.mol^{-1} above A-RNA), is similar to that for the GU-1q9a-1**A/B** pair, i.e., high δ (164°), a slight H2´...H5´ steric clash (2.2 Å), and ε/ζ alteration diminishing the stabilizing effect of the O2´...O2P H-bond.

- **Platforms derived from medium/low-resolution structures**: 2ees, 2qus, 2quw

The lower stability of the conformationally uncharacterized GU-2qus-1 (U) system is partially due to a clash of H4´...H5´(n+1) (1.9 Å) caused by the unusual setting of several backbone torsions. In the case of GU-2quw-1 (U), the electrostatic repulsion between the O5´(n+1) and O4´(n+1) lone pairs leads to a pucker transition from C2´-endo to O4´-exo. Although the O5´(n+1)...O4´(n+1) distance remains the same, the latter pucker directs the upper lone pair (the one on the 5´-side of the sugar ring) away from that of O5´. The energetically least stable structure with relative energy 7.0 kcal.mol^{-1} above A-RNA is GU-2ees-1 (with class/suiteness values of III/0.10). Clarification of the energetic penalty here is not straightforward as there are a number of different factors, which in concert render the conformer less stable. This statement also holds for the preceding high-energy systems even though several possible reasons for inferior stabilities (e.g., steric clashes, suboptimal structural parameters of O2´...O2P H-bonds, etc.) have been suggested. Note that the interconnection of the strongly correlated backbone torsion angles often precludes identification of the source of the high energy in the system. Also note that the five outliers of lower intrinsic stability GU(-1q9a-1**B**, -483d-1**B**, -2qus-1, -2quw-1, -2ees-1) either could not be unambiguously assigned to any of the 46 established RNA backbone conformational classes or were assigned to a known class with a low value of confidence, i.e., suiteness. This finding suggests that the initial X-ray structures of these five conformers are rather atypical and might have been resolved with insufficient resolution (likely in the case of GU-2qus-1, GU-2quw-1, GU-2ees-1) or might represent unstable backbone substates (probably in the case of GU-1q9a-1B, GU-483d-1B), which might be compensated by stabilizing interactions elsewhere in the motif.

ApA rSPSOM Outliers - In-depth Analysis

We have tried to analyze the individual structures and to identify some sources of the variability in energy. The magnitude of the standard deviation in the relative energies is predominantly influenced by the AA-1hr2-3 and AA-1hr2-4 systems, both of which are ~ 6 kcal.mol^{-1} above the A-RNA reference structure. The markedly lower stability of AA-1hr2-3 (IV/0.01) is probably due to the occurrence of the ε torsion angle (308°) in the *gauche*$^{-}$ domain, which is in a forbidden RNA region [S19, S20]. In the case of the unassigned AA-1hr2-4 (U) system, the ε torsion is shifted to even higher values (345°) not observed in experimental RNA structures [S19, S20]. The unnaturally high values of the ε torsion angles lead, among other things, to a steric clash between O1P and H4´, which come as close as 2.2 Å and 2.1 Å in AA-1hr2-3 and AA-1hr2-4, respectively. The initial O1P...H4´ distance in AA-1hr2-4 of ~ 1.6 Å induces a strain which (in combination with the constrained δ torsion) is relaxed during optimization via transition to an atypical C1´-endo pucker. The lower stability of the conformationally uncharacterized AA-1hr2-5 (U) system (3.9 kcal.mol^{-1}) could be also partially ascribed to the uncommon value of ε (49°), which initially positions H4´ in proximity with O1P (1.8 Å). Although the H4´...O1P distance is extended throughout constrained optimization to 2.3 Å, some residual steric clash-related energetic penalty likely remains. The repulsion might be balanced, at least to a certain degree, by an unusual stabilizing O2´...O1P H-bond.

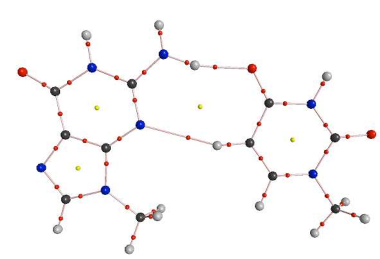
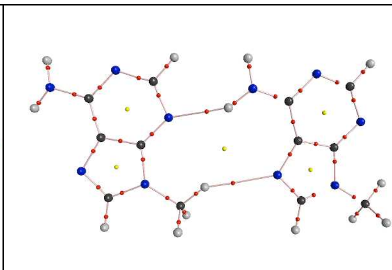
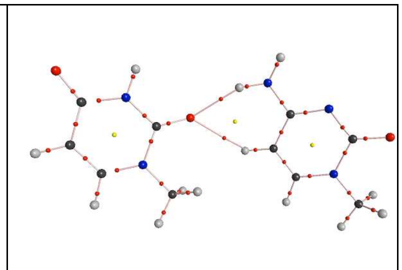
UpC rSPSOM - In-depth Analysis

The lower intrinsic stability of the UC-1drz-1 (U) structure is probably due to the high value of the ε torsion angle (354°), a state rarely observed in high-resolution experimental structures. Note that an ε torsion close to 0° positions the anionic O1P oxygen immediately below the 5´-sugar ring and likely leads to: (i) steric conflicts, e.g., the initial 1.8 Å distance of H4´...O1P in UC-1drz-1 increases in the course of our heavily constrained optimization to 2.1 Å, and (ii) electrostatic repulsion between O1P and O4´. Even though the starting, i.e., experimental, O1P...O4´ distance in the UC-1drz-1 structure increases during minimization from 3.1 Å to 3.5 Å (within the limits imposed by extensive constraints), some residual tension likely remains. The O1P...O4´ distance extension is also accompanied by a C3´-exo → C1´-exo transition. The very low value of the ε torsion angle (7°) in UC-1vc7-1 (U), however, is compensated by the favorable setting of adjacent backbone dihedrals, which prevent the repulsive interactions observed in UC-1drz-1. Despite the unnatural value of the ε torsion (62°) in UC-1u0b-1 (a state incompatible with any of the RNA conformational classes), the structure is partially stabilized by an "outlandish" O2´...O1P H-bond (as in the case of the AA-1hr2-5 rSPSOM). The advantageous setting of the backbone dihedrals in the rSPSOM representation of the UC-1sj3-1 (U) system cooperatively renders this conformer to be of great stability. The fact that the given backbone conformation does not fit into any defined conformational class, however, raises questions about its suitability within a real physiological environment.

[S19] Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; Hershkovits, E.; Williams, L. D.; Keating, K. S.; Pyle, A. M.; Micallef, D.; Westbrook, J.; Berman, H. M. *RNA*. **2008**, *14*, 465-481

[S20] Schneider, B.; Moravek, Z.; Berman, H. M. *Nucleic Acids Res*. **2004**, *32*, 1666-1677

Molecular graphs of GU/AA/UC base pairs Molecular graphs of the most stable GU/AA/UC platform-derived base pairs obtained from the AIM analysis and showing all identified (3, -1) critical points (small red circles) indicative of interatomic interactions and (3, +1) critical points (yellow circles), which give evidence of a circle. The coloring of the atoms follows the standard convention, i.e., oxygen-red, nitrogen-blue, carbon-black, and hydrogen-gray.

		
GU-2qus-1	AA-1hr2-3	UC-1drz-1